Ann. Inst. Statist. Math.
Vol. 49, No. 4, 667–679 (1997)

ESTIMATING DIFFUSION COEFFICIENTS FROM COUNT DATA: EINSTEIN-SMOLUCHOWSKI THEORY REVISITED

N. H. BINGHAM[1] AND BRUCE DUNHAM[2]

[1] *Statistics Department, Birkbeck College (University of London), Malet Street, London WC1E 7HX, U.K.*
[2] *Mathematics Department, University of Nottingham, Nottingham NG7 2RD, U.K.*

(Received November 21, 1995; revised June 27, 1996)

Abstract. The problem of estimating diffusion coefficients has been considered extensively in both discrete and continuous time. We consider here an approach based on counting occupation numbers of diffusing particles. The problem, and our approach, are motivated by statistical mechanics.

Key words and phrases: Diffusion, coverage process, regenerative phenomenon, Campbell's theorem, infinite server queue, Einstein-Smoluchowski process, Avogadro's number.

1. Introduction

The classical Einstein-Smoluchowski theory of diffusion arose in the early years of this century from demonstrations of the reality of atoms and molecules and the experimental determination of Avogadro's number. The work of Einstein (1905, 1906; 1926/56) is reviewed in its historical context by Pais ((1982), Chapter 11); the closely related work by Smoluchowski (1906, 1916) is discussed by Kac ((1985), Chapter 3). Excellent—and now classical—surveys of the theory were given by Chandrasekhar (1943), Kac ((1959), III, §§21–28). As Chandrasekhar ((1943), p. 52) remarks, 'what is perhaps of even greater significance is that we have here the first example of a case in which it has been possible to follow in all its details, both theoretically and experimentally, the transition between the macroscopically irreversible nature of diffusion and the microscopically reversible nature of molecular fluctuations.' Equally striking is Kac's comment ((Kac (1959), pp. 132, 140), on Svedberg's classic data set—517 counts, all in the range from 0 to 6): 'To deduce a number of the order of 10^{23} from Svedberg's numbers none of which exceeded 6 is downright miraculous! Here then is again a result whose worth cannot be judged on its mathematical merits alone.'

Our motivation here is two-fold. First, we are able to revisit the classic treatments of Chadrasekhar (1943) and Kac (1959) armed with the powerful techniques of extensive subsequent developments. Second, the resulting work provides an interesting complement to the sizeable recent literature on statistical inference for stochastic processes, and in particular for diffusion processes, reviewed below. Specifically, we apply our methods, based on count data, to the estimation of the parameters of the Ornstein-Uhlenbeck processes—and, as a limiting case, Brownian motions—arising in the Einstein-Smoluchowski theory.

We turn now to statistical inference for diffusion processes, confining ourselves here to the most basic case, of a diffusion $X = (X_t)$ with drift coefficient b, constant or of known functional form $b(\theta, X_t)$, and diffusion coefficient σ, governed by a stochastic differential equation

$$dX_t = bdt + \sigma dW_t, \qquad X_0 = x$$

with $W = (W_t)$ Brownian motion. The statistical problem of estimating b and σ, given data $(X_s : 0 \leq s \leq t)$ in continuous time, is now classical, and is treated in, for example, Liptser and Shiryaev ((1978), Chapter 17), Kutoyants ((1980/84), Chapter III). We point out, however, that σ is determined in principle, with probability one, from the quadratic variation of X over $[0, t]$ for any $t > 0$; see e.g. Feigin ((1976), Section 5).

The estimation problem in discrete time—with data $(X(t_i))$, where $t_i = i/2^n$ $(0 \leq i \leq 2^n T)$, say—has been much studied recently. See for instance Genon-Catalot *et al.* (1992), Genon-Catalot and Jacod (1993, 1994) and the references cited there, Florens-Zmirou (1993), Brugière (1993).

We consider here a different approach, where the data arise, not from tracking a diffusing particle in time, discrete or continuous, but by counting occupation numbers of a population of diffusing particles within a particular region of observation. This setting originates in statistical physics, where the particles are in suspension in a liquid. As mentioned above, the classical application here of estimating diffusion coefficients—a measure of the mobility of particles—is estimation of Avogadro's number (see Subsection 6.1 for details). A second area where such techniques are classical is biology, where one is to estimate the mobility of spermatazoa, blood cells etc. (Subsection 6.2).

We describe the model—which arises in the classical Einstein-Smoluchowski theory of diffusion—in Section 2 below, summarising the facts we need on the relevant Smoluchowski processes in Section 3. We discuss the statistics of Smoluchowski processes in Section 4, and apply this to the estimation of the relevant Ornstein-Uhlenbeck parameters in Section 5. We close with some remarks and references in Section 6.

2. The model

We consider a suspension of particles in a fluid, in equilibrium. Such a system may be considered at three levels: the microscopic level, of fluid molecules; the mesoscopic level, of diffusing particles (supposed large compared to the fluid molecules but small compared to the region of observation); and the macroscopic level, where a small region within the fluid vessel is kept under observation.

We confine ourselves to the simplest case, the classical model for the Einstein-Smoluchowski theory of diffusion, considered by Chandrasekhar (1943) and Kac

((1959), Chapter III). We assume that the system is in equilibrium, and that particles enter the region A of observation at the instants of a Poisson process, of rate λ say (see e.g. Doob (1953), VIII.5 for background to these assumptions). The mean length of time that a particle spends in A between entering and exiting, α say, is the parameter of principal interest, as $1/\alpha$ is a measure of the *mobility* of the particles. This is related to the mean number μ of particles present in A by Little's formula $\mu = \lambda\alpha$ (see e.g. Mehdi (1991), §2.6). The velocity V_t of a particle is governed by a stochastic differential equation of Ornstein-Uhlenbeck type,

$$dV_t = -\beta V_t dt + c dW_t, \tag{2.1}$$

where β is the drag coefficient and the first term represents frictional forces, the second bombardment by the molecules of the fluid. It is convenient to write

$$c^2 = 2\beta^2 D;$$

then the velocity V_t given V_0 has density (in one dimension for convenience)

$$[2\pi\beta D(1 - e^{-2\beta t})]^{-1/2} \exp\left(-\frac{1}{2}(V - V_0 e^{-\beta t})^2/(\beta D(1 - e^{-2\beta t}))\right),$$

so the limit density as $t \to \infty$ is $N(0, \beta D)$, normal with mean 0 and variance βD. The process is the *Ornstein-Uhlenbeck velocity process* with *diffusion coefficient* D and *relaxation time* $1/\beta$. The limit distribution must coincide with the Maxwell-Boltzmann distribution $N(0, kT/m)$ of statistical mechanics (k is Boltzmann's constant, T the absolute temperature, m the particle's mass), so

$$\beta D = kT/m. \tag{2.2}$$

The covariance is $r(s,t) = D\beta^{-1}(e^{-\beta|t-s|} - e^{-\beta(t+s)})$, whence ($s = t + u$, $s, t \to \infty$) the limiting covariance and correlation are $D\beta^{-1}e^{-\beta|u|}$, $e^{-\beta|u|}$.

Integrating the velocity process to get the Ornstein-Uhlenbeck displacement process $X = (X_t)$, one finds

$$\begin{aligned} EX_t &= x_0 + \beta^{-1}(1 - e^{-\beta t})V_0, \\ \operatorname{var} X_t &= 2Dt + D\beta^{-1}(-3 + 4e^{-\beta t} - e^{-2\beta t}). \end{aligned}$$

For t large compared to the relaxation time $1/\beta$ (typically $1/\beta = O(10^{-8})$ to $O(10^{-10})$ seconds in the statistical physics setting), one has approximately

$$EX_t \sim x_0, \qquad \operatorname{var} X_t \sim 2Dt,$$

and X_t is approximated by a Brownian motion started at X_0 with *diffusion coefficient* D. The approximation

$$\operatorname{var} X_t = 2Dt$$

is the *Einstein relation*; for background, see e.g. Lebowitz and Rost (1994) and the references cited there. For the derivation of the Ornstein-Uhlenbeck process above, see Chandrasekhar ((1943), II.1,2, III.2,3), for the approximation by Brownian motion see Nelson ((1967), §§4, 9), and for the Maxwell-Boltzmann law, see e.g. Gross (1982). A detailed derivation of the Ornstein-Uhlenbeck dynamics at mesoscopic level from dynamic assumptions at microscopic level is given by Dürr *et al.* (1981).

Because individual particles may be difficult to observe directly, or track over time, one may instead observe the number N_t of particles present within the region A at time t. The process $N := (N_t : t \geq 0)$ is called a *Smoluchowski process* in honour of Smoluchowski's pioneering work; in the physics literature, the use of count data (N_t) to study the dynamic parameters D, β (or α, λ) is called *number fluctuation spectroscopy*; see e.g. Brenner *et al.* (1978) for background and references.

3. Smoluchowski processes

We turn to the law of the Smoluchowski process.

PROPOSITION 3.1. *In the Smoluchowski process $N = (N_t)$, N_t is Poisson distributed with mean $\mu = \alpha\lambda$, the mean number of particles in A.*

PROOF. This is an instance of Campbell's theorem (for which see e.g. Moran (1984), Feller (1971)). See Hall ((1988), §2.1), or Takács ((1962), §3.2, Theorem 1) for an alternative approach.

There are two other points of view from which Smoluchowski processes may be considered. The first is that of the *infinite-server queue* $M/G/\infty$ (M for the Markovian character of the input stream, a Poisson process, G for the general service-time distribution, ∞ for the number of servers), when N_t is the 'queue-size', or number of customers being served; see e.g. Takács ((1962), Chapter 3). More generally, the Smoluchowski process is a *shot-noise process*; for background and references, see e.g. Moran (1984), Feller (1971). The second—which concerns us more—is that of *coverage processes*. Here, we work with *reduced data* $I(N_t > 0)$ rather than with count data N_t. That is, we merely record whether or not the region A is occupied by one or more particles at time t. Let ξ_i be the time at which the i-th particle enters A, η_i be the time it spends in A before exiting. The intervals $I_i := (\xi_i, \xi_i + \eta_i)$ are called *segments*. Their union $I := \bigcup_{i=1}^{\infty} I_i$ is called a (simple, linear) *Boolean model*, the most basic type of coverage process. The connected components of I (intervals when A is occupied) are called *clumps*, the connected components of the complement of I (when A is empty) are called *spacings*. A thorough treatment of coverage processes, and in particular the statistical estimation theory of their parameters, is given in Hall (1985, 1988); our treatment of reduced data is based on this. Note that the clumps and spacings are the *busy periods* and *idle periods* in the language of queueing theory. For a treatment of the non-equilibrium case from this viewpoint, see Eick *et al.* (1993).

Working with reduced data $I(N_t > 0)$, it is the clumps and spacings, rather than the segments, that are directly observable. The first task is to relate the distribution $C(\cdot)$ of clump-length with the distribution G of segment-length (recall that G has mean α).

PROPOSITION 3.2. (i) *The mean clump-length γ is given by Smoluchowski's formula*

$$\gamma = EC = (e^{\alpha\lambda} - 1)/\lambda.$$

(ii) *The Laplace-Stieltjes transform $\hat{C}$ of C is given by*

$$\hat{C}(s) := \int_0^\infty e^{-sx} dC(x) = 1 + \frac{s}{\lambda} - \left(\lambda \int_0^\infty e^{-st} \exp(-\lambda \int_0^t (1 - G(x))dx)dt\right)^{-1}. \tag{3.1}$$

(iii) *The variance of clump-length is finite if and only if that of segment-length is finite, and then*

$$\operatorname{var} C = 2e^{\alpha\lambda}\lambda^{-1} \int_0^\infty \left(\exp(\lambda \int_t^\infty (1 - G(x))dx) - 1\right) dt - \frac{(e^{\alpha\lambda} - 1)^2}{\lambda^2}.$$

PROOF. This is Theorem 2.2 of Hall (1988). For (i), (ii), see also Kac ((1959), III.28.10, III.28.8).

Since G has mean α, $\alpha = \int_0^\infty (1 - G(x))dx$, so

$$P(t) := \frac{1}{\alpha} \int_0^t (1 - G(x))dx$$

is a distribution function. Using the language of renewal theory to interpret a particle's lifetime as the time it spends in A, this identifies $P(t)$ as the *residual lifetime law* in equilibrium (or, at great age),

$$P(t) = P(\text{a given particle will have left } A \text{ by time } t \mid \text{the particle is in } A \text{ at time } 0)$$

(see e.g. Feller (1971), VI.7, XI.3). Here $P(t)$ (or rather $1 - P(t)$) is Smoluchowski's key concept, his *probability after-effect* or *Wahrscheinlichkeitsnachwirkung*. Note that

$$P'(0) = 1/\alpha,$$

our measure of particle mobility.

The function $P(t)$, which relates to an individual particle, is closely related to another function, $p(t)$, which refers to the region A under observation:

$$p(t) := P(N_t = 0 \mid N_0 = 0)$$

(because the system is in equilibrium, the model is stationary under time-shifts, so also $p(t) = P(N_{t+s} = 0 \mid N_s = 0)$). For the next result, recall the Kingman theory of regenerative phenomena (Kingman (1964, 1972)).

PROPOSITION 3.3. (i) *The epochs of emptiness of* A, $(I(N_t = 0) : t \geq 0)$, *form a regenerative phenomenon in Kingman's sense, with Kingman p-function* $p(t)$.

(ii) *The functions* $p(t)$, $P(t)$ *are linked by*

$$p(t) = \exp(-\lambda \int_0^t (1 - G(x))dx) = \exp(-\lambda \int_0^\infty \min(t,x)dG(x)) = \exp(-\lambda\alpha P(t)).$$

(iii) *The right-derivative* $p'(0) = p'(0+)$ *of* $p(t)$ *at the origin exists, and* $-p'(0) = \lambda$.

PROOF. (i) The form of equation $(*)$ is characteristic of the Kingman theory, and identifies the p-function; the remaining calculations are easy. See e.g. Kingman ((1964), (22), (40)), Kingman ((1970), (3), (4)), Kingman ((1972), (2.5.12)).

The Smoluchowski process (N_t) is not in general Markovian. There are two ways to handle the mathematical difficulties this poses. The first is to pass from count data (N_t) to reduced data $I(N_t = 0)$ and use the regenerative property above at epochs when $N_t = 0$. This essentially uses the lack of memory property of the exponential distribution (Poisson character of the input stream) to give a partial substitute for the Markov property, holding only at state 0.

The second approach is to model N_t by an *emigration-immigration process*, a particular case of a birth-death process with birth-rates $\lambda_n \equiv \lambda$ (reflecting the Poisson input stream of rate λ) and death-rates $\mu_n = n/\alpha$ (reflecting a propensity of each particle present to exit at rate $1/\alpha$, the particle mobility). The process N is then Markovian, and one can estimate λ and α by standard estimation theory for Markov processes (see Subsection 4.B). This model has been considered in some detail by Bartlett ((1978), §§3.41, 6.31, 8.3), who compares it (§5.21) with the general case; see also Chandrasekhar ((1943), §III.3). Mathematically, this model specialises from the $M/G/\infty$ queue to the $M/M/\infty$ queue (this involves an approximation, discussed further in Subsection 6.3). It is interesting to note that Smoluchowski's formula $\gamma = (e^{\alpha\lambda} - 1)/\lambda$ now follows from that for the mean recurrence time of state 0 in a birth and death process (Keilson (1965)).

4. Statistics of Smoluchowski processes

4.A *Discrete time*

Bartlett (1978) gives a discrete-time treatment by maximum-likelihood (§8.3), and also applies time-series methods (§9.13). Lindley (1956) also gives several time-series methods. A detailed discussion of design of experiments in this context is given by Brenner *et al.* (1978), who also give extensive references to the literature and mention numerous applications.

4.B *Continuous time, count data*

We specialise here to the emigration-immigration model of Section 3 ($M/M/\infty$ queue), with N Markovian. The standard estimation theory for birth-and-death processes, for which see e.g. Billingsley ((1961), §7), applies directly to (N_t). For birth-rates λ_i and death-rates μ_i, the log-likelihood is given by

$$\ell_t = \Sigma_{i=0}^{\infty} u_i(t) \log \lambda_i + \Sigma_{i=1}^{\infty} d_i(t) \log \mu_i - \Sigma_{i=0}^{\infty} \gamma_i(t)(\lambda_i + \mu_i),$$

with $u_i(t)$, $d_i(t)$ the number of jumps in $[0,t]$ from state i up to $i+1$ and down to $i-1$, $\gamma_i(t)$ the time spent in state i. Here $\lambda_i = \lambda$, $\mu_i = i/\alpha$, so

$$\ell_t(\alpha, \lambda) = U(t) \log \lambda - D(t) \log \alpha + \Sigma_1^{\infty} d_i(t) \log i - t\lambda - \frac{1}{\alpha} \int_0^t N_u du,$$

where $U(t)$, $D(t)$ are the number of jumps up and down, $\int_0^t N_u du = \Sigma_1^{\infty} i\gamma_i(t)$ the total time spent in A by all particles. Thus $(U(t), D(t), \int_0^t N_u du)$ is a sufficient statistic for (α, λ) (Billingsley (1961), Example 7.2). The maximum-likelihood estimators are

$$(1/\hat{\alpha})(t) = D(t) / \int_0^t N_u du, \qquad \hat{\lambda}(t) = U(t)/t,$$

the *occurrence-exposure rates*. For background and references, see e.g. Keiding (1975), Daley and Vere-Jones ((1988), §13.3).

Both estimators are consistent and asymptotically normal with rate $\sqrt{t}$ (Billingsley (1961)). Note that $D(t)$, $U(t)$ and $\int_0^t N_u du$ grow at rate t a.s., whereas $U(t) - D(t)$, the change in the number of particles in A, is likely to remain small as the process is in equilibrium. Thus the sample mean

$$\hat{\mu}(t) = \frac{1}{t} \int_0^t N_u du,$$

used to estimate $\mu = \alpha\lambda$, differs little from $\hat{\lambda}(t)/(1/\hat{\alpha})(t)$. We may (and shall) thus regard λ and μ, as second parameters subsidiary to α, as equivalent. It is interesting to note that, since the input stream is Poisson with rate λ, so is the output stream (for this and other properties of queueing output processes, see e.g. Daley (1976), or Asmussen (1987), III.4). We could thus use $\tilde{\lambda}(t) := D(t)/t$ as an alternative estimator for λ, and then $\hat{\mu}(t) = \tilde{\lambda}(t)/(1/\hat{\alpha})(t)$, reflecting $\mu = \lambda\alpha$ more simply.

4.C *Continuous time, reduced data*

For convenience, write $\Lambda := 1/\lambda$. Our reduced data $I(N_t = 0)$ give us

(i) lengths $E_1, E_2, \ldots$ of successive periods of emptiness (spacings), exponentially distributed with parameter λ, mean Λ;

(ii) lengths $C_1, C_2, \ldots$ of successive periods of occupation (clumps) separating the E_j, with law C and mean γ.

The E_j and C_j are all independent (Kingman (1964) or (1972)).

By Smoluchowski's formula,

$$\gamma = \frac{1}{\lambda}(e^{\alpha\lambda} - 1) = \Lambda(e^{\alpha/\Lambda} - 1), \qquad \alpha = \Lambda \log\left(\frac{\gamma}{\Lambda} + 1\right).$$

We form the sample means

$$\hat{\Lambda} := \frac{1}{n}\Sigma_1^n E_i, \qquad \hat{\gamma} := \frac{1}{n}\Sigma_1^n C_i.$$

By the central limit theorem,

$$\sqrt{n}(\hat{\Lambda} - \Lambda) \to N(0, \sigma^2(\Lambda)), \qquad \sqrt{n}(\hat{\gamma} - \gamma) \to N(0, \sigma^2(\gamma)) \qquad (n \to \infty),$$

where the limiting variances follow by calculation on the exponential law or by Hall's result, Proposition 3.2(iii).

The parameter of interest, the mobility $1/\alpha$, is not directly observable, but since

$$1/\alpha = f(\Lambda, \gamma) := 1/(\Lambda \log(1 + \gamma/\Lambda)),$$

we form the empirical estimator

$$(1/\hat{\alpha})_n = 1/(\hat{\Lambda}_n \log(1 + \hat{\gamma}_n/\hat{\Lambda}_n)).$$

By first-order Taylor expansion ('delta method'),

$$\sqrt{n}\,((1/\hat{\alpha})_n - (1/\alpha)) \to N(0, \sigma^2(1/\alpha)) \qquad (n \to \infty)$$

where

$$\sigma^2(1/\alpha) - \sigma^2(\Lambda)(\partial f/\partial\Lambda)^2 + \sigma^2(\gamma)(\partial f/\partial\gamma)^2$$

(see e.g. Rao (1973), §62.4). So given a long enough sequence of lengths $E_1, C_1, \ldots, E_n, C_n$, we may estimate the mobility $1/\alpha$ to arbitrary precision.

4.D *Comparison*

The statistical efficiencies of the methods in 4.B for count data and 4.C for reduced data are not directly comparable. In 4.B, the asymptotic variances are functions of the parameters α, λ, and may be calculated by standard estimation theory, as in Billingsley (1961). However, in 4.C the asymptotic variance depends, through Hall's result (Proposition 3.2(iii)) on the whole function $G(x)$ (or $P(t)$), which in turn depends on the details of the geometry of the region A. We return to this comparison in Subsection 6.3 below.

5. Ornstein-Uhlenbeck parameters

There are two aspects to the work described above: the Smoluchowski process N at macroscopic level, with parameters α, λ (or α, μ) estimable as above, and the Ornstein-Uhlenbeck dynamics of the particles at mesoscopic level, with parameters D, β. We turn now to the relations between the two, following Chandrasekhar ((1943), III.1–3). First,

$$\text{var}[N((i+1)\tau) - N(i\tau)] = 2\mu P(\tau) \tag{5.1}$$

(Chandrasekhar (1943), (361)), and

$$P(t) = 1 - \frac{1}{|A|} \int_A \int_A p(t, \boldsymbol{x}, \boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}, \tag{5.2}$$

where $p(t, \boldsymbol{x}, \boldsymbol{y})$ is the transition density of the Ornstein-Uhlenbeck displacement process with parameters D, β (cf. Chandrasekhar (1943), (380)).

We work in one dimension for simplicity. Use of the limiting Maxwell-Boltzmann distribution and a flux argument gives

$$P'(0) = \left(\frac{\sigma}{v}\right) \left(\frac{kT}{2\pi m}\right)^{1/2} = \left(\frac{\sigma}{v}\right) \left(\frac{\beta D}{2\pi}\right)^{1/2}, \tag{5.3}$$

where σ, v are the surface area and volume of the region A of observation (Chandrasekhar (1943), (414)), both assumed known. Thus βD can be estimated from our estimate of $1/\alpha = P'(0)$.

The argument may be carried through much more generally: although the functional form of $P(t)$ requires detailed knowledge of the transition density, the derivative $P'(0)$ does not. Indeed, Lindley (1956) identifies $P'(0)$ with a measure of partial mobility under only weak qualitative assumptions on the velocity density.

The theory above leads naturally to an estimate of βD; to estimate D and β separately requires a different approach, using Brownian rather than Ornstein-Uhlenbeck dynamics. Recall that, for time-scales t large compared to the relaxation-time $1/\beta$, the Ornstein-Uhlenbeck process with parameters D, β approximates the Brownian motion with diffusion parameter D, in a sense made precise in, e.g., Theorem 9.3 of Nelson (1967) (cf. Chandrasekhar (1943), II.2). Of course, this approximation, valid only for $t \gg 1/\beta$, precludes use of $P'(0)$ as above—and, as (4) now becomes

$$P(t) = 1 - \frac{1}{|A|} \int_A \int_A \exp\left(\frac{-|\boldsymbol{x} - \boldsymbol{y}|^2}{4Dt}\right) (4\pi D t)^{-3/2} d\boldsymbol{x} d\boldsymbol{y}, \tag{5.4}$$

one now has $P(t) \sim ct^{1/2}$ as $t \downarrow 0$, so $P'(0) = \infty$ (Chandrasekhar (1943), (403)). Working in one dimension for simplicity, with h the width of A, one may perform the integrations in (5.4), obtaining

$$1 - P(t) \sim \frac{h}{2\sqrt{\pi D t}} \qquad (t \to \infty).$$

Since $P(t)$ may be estimated from the sample variance of counts $N(it)$ at interval t by (5.1), the diffusion coefficient D may be estimated from this, and hence β from our previous estimate of $D\beta$.

6. Remarks

6.1 *Estimation of Avogadro's number*

The classic motivation for the Einstein-Smoluchowski theory mentioned above, and discussed in detail in Chandrasekhar (1943), was the experimental estimation of Avogadro's number. The drag coefficient β in (2.1) is given by Stokes' law:

$$\beta = 6\pi a\eta/m,$$

with a the particle radius, m the particle mass, η the viscosity of the fluid (for background and restrictions, like flow of low Reynolds number, particles very large compared to fluid molecules, etc., see e.g. Lamb (1932/93), §342). Now $\beta D = kT/m$, where the Boltzmann constant k is R/N, with N Avogadro's number and R the universal gas constant (the constant in Boyle's law $PV = RT$: see e.g. Flügge (1959), p. 25). Thus (2.2) is

$$\beta D = \frac{RT}{Nm}. \tag{6.1}$$

Assuming R known from Boyle's law and T, m measured, Avogadro's number N may be estimated from an estimate of βD. Thus methods such as those of Section 4, which lead to estimates of βD, will provide more accurate means of estimating N than the classical methods of Section 5, which lead to estimates of D. As we remarked in the Introduction, this was the original motivation for this study.

Of course, much more accurate methods of measuring Avogadro's number are available nowadays. See e.g. Seyfried *et al.* (1992) for an X-ray diffraction method accurate to 1.1 parts per million.

6.2 *Applications to biology*

We cite two areas of biological importance.

(a) Mobility of spermatazoa: following the classic experiments by Lord Rothschild (1953), a thorough treatment of Rothschild's problem along the lines of Smoluchowski's work was given by Lindley. For subsequent theoretical developments, see Ruben (1963).

(b) Mobility of leukocytes: a detailed study of movement of white blood cells is given by Brenner *et al.* (1978), who also give further references.

6.3 *The exponential service-time assumption*

We return to the comparison between the results of the $M/G/\infty$ analysis of Subsection 4.C and the $M/M/\infty$ analysis of Subsection 4.B, in the light of Section 5.

The 'service-time distribution'—distribution of the length of time a particle spends in A between entering and its first subsequent exit - clearly depends on the details of the geometry of A. [The length of time spent in A also depends, of course, on the velocity of entry, so one should average this over the limiting velocity law, the Maxwell-Boltzmann law, in equilibrium.] Finding this law in general is an intractable problem, and so the simplifying '$M/M/\infty$' assumption—that the law is exponential—is useful as a general approximation, as in Bartlett ((1978), §5.21).

The question remains, however, of deriving this law—for simplicity, in the simplest case, of one dimension with A an interval $[a, b]$—from the Ornstein-Uhlenbeck dynamics of Section 5. We leave this question open here, referring only to the remarks in Rogers and Williams ((1994), p. 54) on the intractability of the (non-Markov) Ornstein-Uhlenbeck displacement process, and citing Doering *et al.* (1989*a*, 1989*b*), Hesse (1991) for relevant approximations. However, this process may be approximated by a Brownian motion, as in Section 5, for time-scales long compared to the relaxation time. The Brownian analogue of the problem above degenerates (because particles exist instantaneously on entering $[a, b]$). However, the exit-time law of a particle started within (a, b) is known; see e.g. Rogers and Williams ((1994), I, (9.3), (9.4)), Feller ((1971), X, (5.9)); the solution involves the Jacobi theta-functions. Use of Poisson's summation formula gives a rapidly-convergent series expansion from which the tail-behaviour for large time can be read off. This is approximately exponential, which provides support for the applicability of the $M/M/\infty$ approximation.

The questions raised above motivate current work on approximations related to the exponential service-time assumption by the first author and Pitts. We close by noting the interesting fact (pointed out to us by a referee) that Milne (1970) showed that, given λ, the service-time distribution G is identifiable from a realization of each of the input and output processes.

Acknowledgements

We thank Susan Pitts, Aimé Lachal and the referees for their comments.

References

Asmussen, S. (1987). *Applied Probability and Queues*, Wiley, Chichester.

Bartlett, M. S. (1955/78). *An Introduction to Stochastic Processes, with Special Reference to Methods and Applications*, Cambridge University Press, Cambridge (1st ed. 1955, 2nd ed. 1966, 3rd ed. 1978).

Billingsley, P. (1961). *Statistical Inference for Markov Processes*, University Chicago Press, Chicago.

Bingham, N. H. and Pitts, S. M. (1996). Non-parametric inference for the $M/G/\infty$ queue, Statistics Reports 96.8, Birkbeck College, London (to appear).

Brenner, S. L., Nossal, R. J. and Weiss, G. H. (1978). Number fluctuation analysis of random locomotion: statistics of a Smoluchowski process, *J. Statist. Phys.*, **18**, 1–18.

Brugière, P. (1993). Théorème de limite centrale pour un estimateur non-paramétrique de la variance d'un processus de diffusion multidimensionnelle, *Ann. Inst. H. Poincaré Probab. Statist.*, **29**, 357–389.

Chandrasekhar, S. (1943). Stochastic processes in physics and astronomy, *Rev. Modern Phys.*, **15**, 1–89 (reprinted in Wax, N. (ed.) (1954): *Selected Papers on Noise and Stochastic Processes*, Dover, New York, and *Selected papers of S. Chandrasekhar*, Volume 3 (1989), University Chicago Press, Chicago).

Daley, D. J. (1976). Queueing output processes, *Advances in Applied Probability*, **8**, 395–415.

Daley, D. J. and Vere-Jones, D. (1988). *An Introduction to the Theory of Point Processes*, Springer, New York.

Doering, C. R., Hagan, P. S. and Levermore, C. D. (1989*a*). Mean exit times for particles driven by weakly coloured noise, *SIAM J. Appl. Math.*, **49**, 1480–1513.

Doering, C. R., Hagan, P. S. and Levermore, C. D. (1989*b*). The distribution of exit times for weakly coloured noise, *J. Statist. Phys.*, **54**, 1321–1351.

Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

Dürr, D., Goldstein, S. and Lebowitz, J. L. (1981). A mechanical model of Brownian motion, *Comm. Math. Phys.*, **78** 507–530.

Eick, S., Massey, W. A. and Whitt, W. (1993). The physics of the $M_t/G/\infty$ queue, *Oper. Res.*, **41**, 731–742.

Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, *Ann. Physik*, **17**, 549–559 (translated as Chapter I in Einstein (1926/56)).

Einstein, A. (1906). Zur Theorie der Brownschen Bewegung, *Ann. Physik*, **19**, 371–381 (translated as Chapter II in Einstein (1926/56)).

Einstein, A. (1926/56). *Investigations on the Theory of the Brownian Movement*, Methuen, London (1st ed., 1926); Dover, New York (2nd ed., 1956).

Feigin, P. D. (1976). Maximum-likelihood estimation for continuous-time stochastic processes, *Advances in Applied Probability*, **8**, 712–736.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Volume 2, 2nd ed., Wiley, New York.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations, *J. Appl. Probab.*, **30**, 790–804.

Flügge, S. (ed.) (1959). *Prinzipien der Thermodynamik und Statistik*, Springer, Handbuch der Physik, Band III/2, Berlin.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, *Ann. Inst. H. Poincaré Probab. Statist.*, **29**, 119–151.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling, *Scand. J. Statist.*, **21**, 193–221.

Genon-Catalot, V., Laredo, C. and Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelets methods, *Scand. J. Statist.*, **19**, 317–335.

Gross, L. (1982). Thermodynamics, statistical mechanics and random fields, Ecole d'Été de Probabilités de Saint-Flour X, *Lecture Notes in Math.*, **929**, 101–204, Springer, Berlin.

Hall, P. (1985). On inference in a one-dimensional mosaic and an $M/G/\infty$ queue, *Advances in Applied Probability*, **17**, 210–229.

Hall, P. (1988). *An Introduction to the Theory of Coverage Processes*, Wiley, New York.

Hesse, C. H. (1991). The one-sided barrier problem for an integrated Ornstein-Uhlenbeck process, *Comm. Statist. Stochastic Models*, **7**, 447–480.

Kac, M. (1959). *Probability and Related Topics in the Physical Sciences*, Interscience, New York.

Kac, M. (1985). *Enigmas of Chance. An Autobiography*, Harper & Row, New York.

Keiding, N. (1975). Maximum likelihood estimation in the birth-and-death process, *Ann. Statist.*, **3**, 363–372 (Correction: *ibid.* (1978). **6**, p. 472).

Keilson, J. (1965). A review of transient behaviour in regular birth-death processes II, *J. Appl. Probab.*, **2**, 405–428.

Kingman, J. F. C. (1964). The stochastic theory of regenerative events, *Zeit. Wahrscheinlichkeitsth.*, **2**, 180–224.

Kingman, J. F. C. (1970). An application of the theory of regenerative phenomena, *Proceedings of the Cambridge Philosophical Society*, **68**, 697–701.

Kingman, J. F. C. (1972). *Regenerative Phenomena*, Wiley, London.

Kutoyants, Yu. A. (1980/84). *Parameter Estimation for Stochastic Processes* (in Russian), *Akad. Nauk Armyan SSR Dokl.* (1980); English translation, *Res. Exp. Math.*, **6** (1984).

Lamb, H. (1932/93). *Hydrodynamics*, 6th ed., Cambridge University Press, Cambridge (1st ed., *Treatise on the Mathematical Theory of the Motion of Fluids*, 1879).

Lebowitz, J. L. and Rost, H. (1994). The Einstein relation for the displacement of a test particle in a random environment, *Stochastic Process Appl.*, **54**, 183–196.

Lindley, D. V. (1956). The estimation of velocity distributions from counts, *Proceedings of the International Congress of Mathematicians* (Amsterdam, 1954) III, 427–444, North-Holland, Amsterdam.

Liptser, R. S. and Shiryaev, A. N. (1978). *Statistics of Random Processes, Volume II: Applications*, Springer, New York.

Mehdi, J. (1991). *Stochastic Models in Queueing Theory*, Academic Press, Boston.

Milne, R. K. (1970). Identifiability for random translates of Poisson processes, *Zeit. Wahrscheinlichkeitsth.*, **15**, 195–201.

Moran, P. A. P. (1968/84). *An Introduction to Probability Theory*, 2nd ed., Oxford University Press, Oxford (1st ed. 1968, 2nd ed. 1984).

Nelson, E. (1967). *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton.

Pais, A. (1982). *'Subtle is the Lord...'. The Science and the Life of Albert Einstein*, Oxford University Press, Oxford.

Rao, C. R. (1973). *Linear Statistical Inference and Applications*, 2nd ed., Wiley, New York.

Rogers, L. C. G. and Williams, D. (1994). *Diffusions, Markov Processes and Martingales, Volume One: Foundations*, 2nd ed., Wiley, Chichester.

Rothschild, Victor, Lord (1953). A new method for estimating the speed of spermatazoa, *Journal of Experimental Biology*, **30**, 178–199.

Ruben, H. (1963). The estimation of a fundamental interaction parameter in an emigration-immigration process, *Ann. Math. Statist.*, **34**, 238–259.

Seyfried, P. *et al.* (1992). A determination of the Avogadro constant, *Zeitschrift für Physik B*, **87** (3), 289–298.

Smoluchowski, M. von (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen, *Ann. Physik*, **21**, 756–780.

Smoluchowski, M. von (1916). Drei Vorträge über Diffusion, Brownschen Bewegung und Koagulation von Kolloidteilchen, *Physikalische Zeitschrift*, **17**, 557–571 and 587–599.

Takács, L. (1962). *Introduction to the Theory of Queues*, Oxford University Press, Oxford.